March, 1966

FIDUCIAL THEORY AND INVARIANT PREDICTION[1]

By R. B. Hora[2] and R. J. Buehler

Technical Report No. 73

University of Minnesota
Minneapolis, Minnesota

[1] Research supported by National Science Foundation Grants G19126 and GP3816.

[2] Now at Purdue University, Indianapolis, Indiana.

FIDUCIAL THEORY AND INVARIANT PREDICTION[1]

by R. B. Hora[2] and R. J. Buehler

1. Introduction. We are concerned with the prediction of future observations y and functions $\psi(y)$ given past observations x when x and y have a joint distribution of known form which depends on an unknown parameter ω. The problem of prediction has many similarities with that of estimation, and as in estimation, several approaches are possible: (i) Decision-theoretic: When ω is the true parameter value, $\psi(y)$ is the random value to be predicted, $\hat{\psi}(x)$ is the predicted value, let the "loss" be given by $L(\psi(y), \hat{\psi}(x), \omega)$, and choose $\psi(\cdot)$ so that the expected loss is small in some sense. The special case $L = (\hat{\psi}-\psi)^2$ gives a mean-square-error criterion. (ii) Bayesian: Our knowledge of y is given by the conditional distribution of y given x where the joint distribution of (x,y,ω) is the product of the prior distribution of ω and the likelihood of (x,y) given ω; the "best" predictor $\hat{\psi}(x)$ minimizes the loss averaged over the joint distribution of (x,y,ω). (iii) Confidence: Find a function $\bar{\psi}(x,\gamma)$ such that $P\{\psi(y) \leq \bar{\psi}(x,\gamma)|\omega\} = \gamma$ for all ω. The limits $\bar{\psi}(x,\gamma_1)$ and $\bar{\psi}(x,\gamma_2)$ then give a prediction analogue of confidence limits for $\psi(y)$ corresponding to confidence level $\gamma_2-\gamma_1$. (iv) Fiducial: Ignorance of ω and knowledge of x is equivalent to knowledge that y belongs to some distribution depending on x but not on ω.

The present paper establishes some connections between the above approaches when the joint distribution of x and y possesses certain

---

[1]Research supported by National Science Foundation Grants G19126 and GP3816.

[2]Now at Purdue University, Indianapolis, Indiana.

invariance properties. A simple example will serve to indicate the nature of the results. Suppose x and y have independent $N(\theta,1)$ distributions (normal with mean θ and unit variance). Under the transformation g_α defined by $g_\alpha(x,y,\theta) = (g_\alpha x,\ g_\alpha y,\ g_\alpha \theta) = (x-\alpha, y-\alpha, \theta-\alpha)$ one has the invariance condition that $g_\alpha x$ and $g_\alpha y$ have independent distributions $N(g_\alpha \theta,1)$. To obtain the "fiducial" distribution of y given x we argue by analogy with Fisher (1935) and (1956), Chapter 5, Section 3, as follows: y-x has the distribution $N(0,2)$ which is independent of θ, so that holding x fixed, by a "pivotal" argument, the fiducial distribution of y given x is $N(x,2)$. We now list several properties of this fiducial distribution, each of which can be easily verified and each of which is generalized in the present paper. (a) The fiducial distribution is identical with that given by Bayesian analysis based on a uniform prior distribution of θ over $(-\infty,\infty)$. (b) Fiducial limits obtained from the fiducial distribution have the frequency interpretation associated with the prediction analogues of confidence limits, for example, $P\{y \leq x + 1.96\sqrt{2}\,|\,\theta\} = 0.975$ for all θ. (c) The predictor $\hat{y}(x)$ which equals the mean of the fiducial distribution, $\hat{y}(x) = x$, minimizes $E(\hat{y}(x) - y)^2$ amongst the class of invariant predictors satisfying $y(x-\alpha) = y(x) - \alpha$.

We conclude this section by indicating the relation between the present paper and earlier work on related problems.

Prediction analogues of confidence limits were discussed in the case of regression models by Eisenhart (1939), and later, for example, by Mood (1950), Section 13.3. In the interest of simplicity, the present paper does not treat the case of a regression structure, which we would expect to be a fairly straightforward extension.

Weiss (1955) gave a general method of determining "confidence sets" for future observations y using a sufficient statistic $T(x,y)$ for ω. Our construction in Section 8 below is similar to his; his method would also apply in certain cases lacking the group structure which we assume.

The Bayesian analysis of the prediction problem has been discussed for example by Fisher (1956), Chapter 3, Section 2, and by Jeffreys (1961), Chapter 3.

In the present paper attention is restricted to continuous variates. Similar problems in the prediction of discrete variates have been considered by Roy (1960) and Thatcher (1964).

Kitagawa (1951) considered estimators α_1^* and α_2^* of a parameter α, based respectively on a past and a future sample and considered the accuracy of prediction of α_2^*. Later Kitagawa (1957) gave a theory of fiducial prediction quite close in spirit to the present work, but depending heavily on the theory of exponential families of distributions and on sufficient statistics, which are not required in the present treatment. Kitagawa's (1957) statistic h in Definition 2.3 and in equations (4.01) and (4.02) is an ancillary function of two sufficient statistics, and plays the same role as the quantity $t^{-1}y$ in the Appendix below. Kudo (1956) applied Kitagawa's theory in obtaining the fiducial distribution of the maximum of a future sample from a normal population. This case falls within the scope of Section 8 below since $\psi(y) = \max(y_1,\ldots,y_m)$ is an invariantly predictable function.

2. Distributional assumptions. In Hora and Buehler (1966) (hereafter referred to as HB-1) five distributional assumptions were given in order to define the fiducial distribution of the parameter ω. Although these assumptions are not identical with Fraser's (1961), they are essentially equivalent. Below we give eight assumptions of which the first five are the same as those of HB-1 and thus assure the existence of a fiducial distribution of the parameter in the sense of Fraser. Assumptions 6, 7 and 8 relate to the conditional distribution of the future observations y given x and ω. The form of the assumptions was chosen primarily for ease of application, and additional discussion my help to clarify the nature of the model. In essence, we have two spaces $\mathcal{X}$ and $\mathcal{Y}$, each of which is transformed onto itself by a group $\mathcal{G}$. A single distribution P on $\mathcal{X}\times\mathcal{Y}$ generates a family of distributions P^g defined by $P^g(gX\times gY) = P(X\times Y)$ $(g\in\mathcal{G},\ X\subset\mathcal{X},\ Y\subset\mathcal{Y})$, so that the group element is identified with the parameter of the distribution. Beyond this, the assumptions are however not symmetric with respect to the spaces $\mathcal{X}$ and $\mathcal{Y}$. While the orbit $\mathcal{G}x = \{gx|g\in\mathcal{G}\}$ of any $x\in\mathcal{X}$ must be in one-to-one correspondence with $\mathcal{G}$, the same is not true of the orbits $\mathcal{G}y$. For example, if $\mathcal{G}$ changes location and scale (Example 3.2 below), $\mathcal{X}$ must be at least two-dimensional, while $\mathcal{Y}$ may be the real line, a "smaller" space than $\mathcal{G}$.

Assumption 1. $(\mathcal{X},B_X)$, $(\mathcal{T},B_T)$, $(\mathcal{A},B_A)$ and (Ω,B_Ω) are measurable spaces such that there is a one-to-one correspondence between $\mathcal{X}$ and $\mathcal{T}\times\mathcal{A}$,

(2.1) $$x = (t,a),$$

and B_X corresponds to the minimal σ-field on $\mathcal{T}\times\mathcal{A}$ generated by B_T and B_A.

Assumption 2. $\mathcal{G} = \{g\}$ is a group and $(\mathcal{G},B_G)$ is a measurable space on which there exists a left invariant Haar measure μ satisfying

(2.2) $$\mu(gG) = \mu(G) \qquad \text{all } g\in\mathcal{G},\quad G\in B_G.$$

Assumption 3. There exist one-to-one correspondences between the three spaces $\mathcal{J}$, Ω and $\mathcal{G}$ such that all images of measurable sets are measurable.

Assumption 4. There is a family P^{ω}, $\omega\epsilon\Omega$, of probability measures on $\mathcal{X}$ such that for corresponding $g\epsilon\mathcal{G}$ and $\omega\epsilon\Omega$

$$(2.3) \qquad P^{\omega}(X) = \int_X f_1(g^{-1}t|a)\lambda(da)\mu(dt) \quad \text{all } X\epsilon B_X$$

where λ is a probability measure on $\mathcal{A}$ and $f_1(\cdot|a)$ is a density with respect to μ on $\mathcal{J}$ for each $a\epsilon\mathcal{A}$.

Here and later we use the notational convention of HB-1 wherein the same letter may be used to denote points in $\mathcal{J}$, Ω or $\mathcal{G}$ which correspond by Assumption 3.

Assumption 5. If ω_1 and ω_2 are distinct, then P^{ω_1} and P^{ω_2} are not identical.

Assumption 6. $(\mathcal{Y}, B_Y)$ is a measurable space, and for each $g\epsilon\mathcal{G}$, gy $(y\epsilon\mathcal{Y})$ is a measurable one-to-one transformation of $\mathcal{Y}$ onto itself.

Assumption 7. ξ is a measure on $\mathcal{Y}$ such that for each $g\epsilon\mathcal{G}$, $Y\epsilon B_Y$, $\xi(Y) = J(g)\xi(gY)$, that is, the Radon-Nikodym derivative $J(g) = \xi(dy)/\xi(g(dy))$ exists and does not depend on y.

Assumption 8. For each $\omega\epsilon\Omega$ there is a probability distribution on $\mathcal{X}\times\mathcal{Y}$ such that the conditional distribution on $\mathcal{Y}$ given x and ω has the form

$$(2.4) \qquad P^{\omega}(Y|x) = \int_Y f_2(\omega^{-1}y|\omega^{-1}t,a)J(\omega)\xi(dy).$$

3. <u>Examples</u>. In HB-1 four examples of location and scale parameter families are given ranging in generality from the case of one location parameter (θ) to the case of two location and two scale parameters $(\theta_1,\theta_2,\sigma_1,\sigma_2)$. In the present section we extend these examples to include future observations y and indicate why the new Assumptions 6, 7 and 8 are satisfied. The notation will correspond with that of HB-1 except for modifications which allow y to be reserved for future observations and m for the number of future observations as indicated in Table 3.1. Thus the space $(\mathcal{Y}, B_Y)$ of Assumption 6 is (R_k, B_k) where R_k is k-dimensional euclidean space, B_k is the class of Borel sets, $k = m$ in Examples 3.1 and 3.2, and $k = m_1 + m_2$ in Examples 3.3 and 3.4.

The transformation of the $\mathcal{Y}$ space is the same as the transformation of the $\mathcal{X}$ space described in HB-1; thus in Example 3.1, $gx = (x_1 + \alpha,\ldots, x_n + \alpha)$, $gy = (y_1 + \alpha,\ldots, y_m + \alpha)$ (or $gy_j = y_j + \alpha$); similarly $gy_j = \alpha + \beta y_j$ in Example 3.2, and $gy_{ij} = \alpha_i + \beta y_{ij}$ and $\alpha_i + \beta_i y_{ij}$ in Examples 3.3 and 3.4 respectively. The measure ξ of Assumption 7 is Lebesgue measure L_k, and $J(g) = 1$, β^{-m}, $\beta^{-m_1-m_2}$, and $\beta_1^{-m_1}\beta_2^{-m_2}$ respectively in the four examples. The form of the conditional density $f_2(\cdot|\cdot,\cdot)$ of Assumption 8 arises when x and y are given arbitrary densities with respect to Lebesgue measure and the family P^{ω} on $\mathcal{X}\times\mathcal{Y}$ is generated by transforming the $\mathcal{X}\times\mathcal{Y}$ space by the element $g\epsilon\mathcal{G}$ corresponding by Assumption 3 to the element $\omega\epsilon\Omega$. More precisely, if P is the arbitrarily given measure then P^{ω} is defined by $P^{\omega}(gX \times gY) = P(X \times Y)$ where ω and g correspond (for example $\omega = (\theta,\sigma)$ corresponds to $g = (\alpha,\beta)$ when $\theta = \alpha$ and $\sigma = \beta$).

Table 3.1

<table>
<tr><th>Example</th><th>ω</th><th>x</th><th>y</th></tr>
<tr><td>3.1</td><td>θ</td><td rowspan="2">$x_1,\ldots, x_n$</td><td rowspan="2">$y_1,\ldots, y_m$</td></tr>
<tr><td>3.2</td><td>θ,σ</td></tr>
<tr><td>3.3</td><td>θ_1,θ_2,σ</td><td rowspan="2">$x_{11},\ldots,x_{1n_1}, x_{21},\ldots, x_{2n_2}$</td><td rowspan="2">$y_{11},\ldots,y_{1m_1},y_{21},\ldots y_{2m_2}$</td></tr>
<tr><td>3.4</td><td>$\theta_1,\theta_2,\sigma_1,\sigma_2$</td></tr>
</table>

4. Bayesian and fiducial distributions. Let ν denote the right Haar measure and Δ the modular function defined by

$$(4.1) \qquad \nu(G) = \mu(G^{-1}), \quad \mu(Gg) = \Delta(g)\mu(G) \quad \text{all } g \in \mathcal{G}, \quad G \in B_G.$$

As is known, the measure ν is unbounded in most cases of interest (we then call it a "quasi-distribution"), but nevertheless is the natural prior measure to use in Bayesian analysis in order to obtain solutions exhibiting invariance properties, as has been noted for example in Peisakoff (1950), Barnard (1952), Hartigan (1964) and in HB-1. The product of the probability elements in (2.3) and (2.4) gives the probability element for the joint distribution of (t,a,y) given ω; and an additional factor of $\nu(d\omega)$ gives the (quasi-) distribution of (t,a,y,ω) in the form

$$(4.2) \qquad f(\omega^{-1}t,\omega^{-1}y|a)\ J(\omega)\lambda(da)\mu(dt)\xi(dy)\nu(d\omega),$$

where, for notational convenience, we have defined $f(\cdot,\cdot|\cdot)$ by

$$(4.3) \qquad f(t,y|a) = f_1(t|a)f_2(y|t,a).$$

Three conditional probability elements derivable from (4.2) are:

$$(4.4) \qquad \text{for } y,\omega|t,a: \quad f(\omega^{-1}t,\omega^{-1}y|a)J(\omega)\Delta(t)\xi(dy)\nu(d\omega)$$

$$(4.5) \qquad \text{for } \omega|t,a: \quad f_1(\omega^{-1}t|a)\Delta(t)\nu(d\omega)$$

$$(4.6) \qquad \text{for } y|t,a: \quad \Delta(t)\xi(dy)\int_\Omega f(\omega^{-1}t,\omega^{-1}y|a)J(\omega)\nu(d\omega).$$

Here (4.4) is obtained from the quotient of (4.2) by its integral over y and ω. The integral of $f_2(\omega^{-1}y|\omega^{-1}t,a)J(\omega)\xi(dy)$ over y is unity; the integral of $f_1(\omega^{-1}t|a)\nu(d\omega)$ over ω is $1/\Delta(t)$ (by changing variables to integrate with respect to μ rather than ν). Thus (4.4) follows, and (4.5) and (4.6) are obtained from it by integrating over y and ω respectively.

From the Bayesian viewpoint (4.4), (4.5), and (4.6) represent the posterior distributions, given past observations $x = (t,a)$, of (ω,y), ω and y

respectively, where ω is the parameter and y denotes future observations.

From the fiducial viewpoint, (4.5) is the usual fiducial distribution of the parameter in the sense of Fraser (1961), as described in the present notation in HB-1.

The expression (4.6) will be called the fiducial distribution of the future observations y given past observations x, and it is to be interpreted as representing the state of our knowledge of y when x has been observed and when nothing else is known about the parameter ω. Two justifications for the terminology are: (1) consistency with Fisher (1935), (1956), pp 113-116, in special cases which are straightforward to verify; and (2) an alternative derivation of (4.6) given in the Appendix, by a pivotal argument which is Fisherian in spirit. Previous generalizations of Fisher's examples have been given by Kitagawa (1957), Sprott (1963), and Ramsey (1963). The present expression is claimed to be more general in that it does not require the existence of sufficient statistics (Kitagawa), nor is it univariate in nature (Sprott), nor is it restricted to location and scale parameter families (Ramsey).

The expression (4.4) will be called the joint fiducial distribution of y and ω given x, and represents the state of our knowledge jointly about y and ω based solely on the observed x (and the assumed probability law). The relationship to (4.5) and (4.6) is the justification for the terminology. Special cases relating to location and scale parameter families have been given previously by Ramsey (1963).

5. An expectation identity. In HB-1 an expectation identity was used to relate fiducial and confidence limits and to give explicit expressions for "best" estimators. The theorem below will be seen to have similar applications to prediction problems in the following sections. The earlier identity would follow from the present one by integrating over y.

Theorem 5.1. If Assumptions 1 through 8 are satisfied and if

$$H(gx,gy,g\omega) = H(x,y,\omega) \tag{5.1}$$

then

$$E^{y,t|a,\omega}H(x,y,\omega) = E_f^{y,\omega|x}H(x,y,\omega) \tag{5.2}$$

where $E^{y,t|a,\omega}$ denotes conditional expectation with respect to (y,t) given (a,ω) and where $E_f^{y,\omega|x}$ denotes expectation with respect to the fiducial distribution of (y,ω) given $x = (t,a)$ (given by (4.4)).

Proof. Defining $H(t,a,y,\omega) = H(x,y,\omega)$ gives $H(gt,a,gy,g\omega) = H(t,a,y,\omega)$. Let e denote the identity. The conditional distribution of (y,t) given (a,ω) is obtained by deleting $\lambda(da)\nu(d\omega)$ from (4.2), so that:

$$\begin{aligned}
E^{y,t|a,\omega}H(x,y,\omega) &= J(\omega)\iint H(t,a,y,\omega)f(\omega^{-1}t,\omega^{-1}y|a)\mu(dt)\xi(dy)\\
&= J(\omega)\iint H(\omega^{-1}t,a,\omega^{-1}y,e)f(\omega^{-1}t,\omega^{-1}y|a)\mu(dt)\xi(dy)\\
&= \iint H(r,a,y',e)f(r,y'|a)\mu(dr)\xi(dy') \quad (r=\omega^{-1}t, y'=\omega^{-1}y)\\
&= \iint H(\omega^{-1}t,a,y',e)f(\omega^{-1}t,y'|a)\Delta(t)\nu(d\omega)\xi(dy') \quad (r=\omega^{-1}t)\\
&= \Delta(t)\iint H(\omega^{-1}t,a,y',e)f(\omega^{-1}t,y'|a)\xi(dy')\nu(d\omega)\\
&= \Delta(t)\iint H(\omega^{-1}t,a,\omega^{-1}y,e)f(\omega^{-1}t,\omega^{-1}y|a)J(\omega)\xi(dy)\nu(d\omega)\\
&= \Delta(t)\iint H(t,a,y,\omega)f(\omega^{-1}t,\omega^{-1}y|a)J(\omega)\xi(dy)\nu(d\omega)\\
&= E_f^{y,\omega|x}H(x,y,\omega).
\end{aligned}$$

Note that the invariance assumption (5.1) has been used in obtaining the second and penultimate equalities. The variate change from y to y' is needed of course to free the expression of ω before ω is re-introduced as an integration variate. The initial expression depends only on (a,ω) and the final one only on $x = (t,a)$; thus all expressions depend only on the ancillary a and are independent of (ω,t).

6. Invariantly predictable functions. We now consider the problem of prediction of a function $\psi(y)$ of the future observations y. In order to exploit the assumed invariance properties of the family of distributions we find it necessary to restrict attention to "invariantly predictable functions", which are the analogs of the "invariantly estimable functions" defined in HB-1. A function $\psi(y)$ is called "invariantly predictable" if

$$\psi(y_1) = \psi(y_2) \text{ implies } \psi(gy_1) = \psi(gy_2), \text{ all } g \in \mathcal{G}. \tag{6.1}$$

As in HB-1 it can be shown that the transformations $\mathcal{G}' = \{g'\}$ of $\{\psi\}$ defined by

$$g'\psi(y) = \psi(gy) \tag{6.2}$$

form a group homomorphic to $\mathcal{G}$.

The relation between ancillary statistics and orbits has been noted for example in HB-1. If we define a "y-related ancillary" as any function b(y) which is constant on each orbit $\mathcal{G}y = \{gy \mid g \in \mathcal{G}\}$, then b(y) is invariantly predictable. In a sense this is however a degenerate case in that the distribution of b(y) is independent of ω, so that predictions would simply be made from the known fixed distribution without using the observations x.

Other examples not having this degeneracy are given in Table 6.1. Q and R denote quadratic functions of the form

$$\begin{cases} Q = \Sigma\Sigma a_{ij} y_k y_j, & R = \Sigma\Sigma a_{ij} y_{1i} y_{2j}, \\ \text{where } \Sigma_i a_{ij} = 0 \text{ (all } j), & \Sigma_j a_{ij} = 0 \text{ (all } i), \end{cases} \tag{6.3}$$

of which the sample variance and covariance, respectively, are examples. Q_1 and Q_2 have the same definition as Q with y_i replaced by y_{1i} and y_{2i} respectively. In addition to the examples we may note that any one-to-one function of an invariantly predictable function is invariantly predictable.

Table 6.1

Example	Invariantly predictable	Not invariantly predictable
3.1	$\Sigma c_i y_i$; $\Sigma c_i y_{[i]}$ (where $y_{[i]}$ is the ith order statistic).	y_1^2; $y_1 y_2$.
3.2	As above; Q (see (6.3)); homogeneous functions of Q.	As above; y_1/Q.
3.3	$\Sigma\Sigma c_{ij} y_{ij}$; R (see (6.3)).	y^2_{11}; $y_{11}y_{21}$.
3.4	$Q_1^r Q_2^s$ (see text); R.	$y_{11} - y_{21}$.

In HB-1 it was shown how the parametric function $\psi(\omega)$ could be used to define two subgroups H and K of $\mathcal{G}$. In the present case where $\psi(y)$ is defined for $y \epsilon \mathcal{Y}$, the analog of H apparently does not exist in general because no one-to-one correspondence need exist between $\mathcal{G}$ and $\mathcal{Y}$. However K can again be defined as the kernel of the homomorphism between $\mathcal{G}$ and $\mathcal{G}'$:

(6.4) $$K = \{g \mid \psi(gy) = \psi(y), \text{ all } y \epsilon \mathcal{Y}\}.$$

If for example $\psi(y) = y_1$, then $\mathcal{G}'$ is exactly transitive on $\{\psi\}$ in Example 3.1, so that a one-to-one correspondence is defined between $\{\psi\}$ and $\mathcal{G}'$. The same would not be true in Example 3.2 where $\mathcal{G}'$ is "larger" than $\{\psi\}$. The former case typifies a general class wherein the values of ψ can be identified with elements $g' \epsilon \mathcal{G}'$. Since the quotient group $\mathcal{G}/K$ having elements gK is known from group theory to be isomorphic with $\mathcal{G}'$, we can in such cases identify the values of ψ with the cosets gK.

7. Best invariant predictors. To treat the prediction problem from a decision theoretic viewpoint (generalized to allow for the "future" observations y) we wish to construct invariant functions on $\mathcal{X} \times \mathcal{Y} \times \Omega$ to represent the "loss" incurred when $\hat{\psi}(x)$ is the predicted value, $\psi(y)$ is the observed value, and ω is the true parameter value. A predictor $\hat{\psi}(x)$ of an invariantly predictable function $\psi(y)$ will be called "invariant" if it satisfies

$$(7.1) \qquad \hat{\psi}(gx) = g'\hat{\psi}(x),$$

where g' is defined in (6.2).

Lemma 7.1. If $\psi(y)$ and $\hat{\psi}(x)$ satisfy (6.1) and (7.1), if $\Phi(\cdot,\cdot)$ is a function on $\Psi \times \Psi$ (where $\Psi = \{\psi\}$), and if

$$(7.2) \qquad H(x,y,\omega) = \Phi(\omega'^{-1}\hat{\psi}(x),\ \omega'^{-1}\psi(y))$$

where $\omega' \in \mathcal{G}'$ is the image of $\omega \in \Omega$ defined in (6.2), then $H(x,y,\omega) = H(gx,\ gy,\ g\omega)$.

Proof. Straightforward.

Theorem 7.1. If (i) Assumptions 1 through 8 are satisfied, (ii) $\psi(y)$ is invariantly predictable, (iii) $\Phi(\cdot,\cdot)$ is a real-valued function on $\Psi \times \Psi$, (iv) there is a unique value $\hat{\psi}_o(x)$ of $\hat{\Psi}$ which minimizes

$$E_f^{y,\omega|x}\Phi(\omega'^{-1}\hat{\psi},\omega'^{-1}\psi(y)),$$

then $\hat{\psi}_o(x)$ minimizes $E^{x,y|\omega}\Phi(\omega'^{-1}\hat{\psi}(x),\omega'^{-1}\psi(y))$ amongst the class of invariant predictors $\hat{\psi}(x)$ satisfying (7.1).

Proof. Similar to that of Theorem 5.1 of HB-1, using Lemma 7.1 and Theorem 5.1 above.

Corollary 7.1. When Ψ is a subset of the real line and when $\omega'^{-1}\psi = \{\varphi(\omega)\}^{1/2}\psi + \lambda(\omega)$, then

(7.4) $$\hat{\psi}_o(x) = E_f\{\varphi(\omega)\psi(y)\}/E_f\varphi(\omega) \qquad (E_f \equiv E_f^{y,\omega|x})$$

is the minimum mean square error invariant predictor of $\psi(y)$, that is, it minimizes $E^{x,y|\omega}\{\hat{\psi}(x) - \psi(y)\}^2$ in the class of invariant predictors.

Proof. Choose $\Phi(u,v) = (u - v)^2$. Then

$$\Phi(\omega'^{-1}\hat{\psi},\omega'^{-1}\psi) = \varphi(\omega)(\hat{\psi} - \psi)^2,$$

and the proof proceeds similarly to that of Corollary 5.1 of HB-1.

Table 7.1 gives several examples wherein Corollary 7.1 applies, and the minimum mean square error invariant (= "best" invariant) predictor is defined in terms of fiducial expectations. In Example 3.4, $\varphi = \sigma_1^{-4r}\sigma_2^{-4s}$.

Table 7.1

Example	$\psi(y)$	$g'\psi(y) = \psi(gy)$	$\Phi(\omega'^{-1}\hat{\psi},\omega'^{-1}\psi)$	"best" invariant predictor
3.1	y_{max}	$\psi(y) + \alpha$	$(\hat{\psi} - \psi)^2$	$E_f\psi$
3.2	$\Sigma a_i y_i$	$\beta\psi(y) + \alpha\Sigma a_i$	$(\hat{\psi} - \psi)^2/\sigma^2$	$E_f(\sigma^{-2}\psi)/E_f\sigma^{-2}$
3.2	Q	$\beta^2\psi(y)$	$(\hat{\psi} - \psi)^2/\sigma^4$	$E_f(\sigma^{-4}\psi)/E_f\sigma^{-4}$
3.3	$\Sigma\Sigma a_{ij}y_{ij}$	$\beta\psi + \Sigma\Sigma\alpha_i a_{ij}$	$(\hat{\psi} - \psi)^2/\sigma^2$	$E_f(\sigma^{-2}\psi)/E_f\sigma^{-2}$
3.4	$Q_1^r Q_2^s$	$\beta_1^{2r}\beta_2^{2s}\psi$	$(\hat{\psi} - \psi)^2\varphi$	$E_f(\varphi\psi)/E_f\,\varphi$

8. Fiducial prediction limits. In this section the prediction analogues of confidence limits already mentioned in Section 1 are considered. It is shown that invariant estimability of $\psi(y)$ essentially ensures that fiducial limits will have a particular frequency interpretation.

For any real valued $\psi(y)$, any observed x, and any probability level γ, an upper fiducial prediction limit $\overline{\psi}(x,\gamma)$ is defined by

$$(8.1)\qquad P_f\{\psi(y) \leq \overline{\psi}(x,\gamma)|x\} = \gamma$$

where P_f denotes fiducial probability. In the usual, reasonably well behaved cases $\overline{\psi}(x,\gamma)$ will exist uniquely. The fiducial prediction limits may or may not possess the frequency interpretation expressed by

$$(8.2)\qquad P\{\psi(y) \leq \overline{\psi}(x,\gamma)|\omega\} = \gamma \text{ for all } \omega.$$

Theorem 8.1. If (i) Assumptions 1 through 8 are satisfied, (ii) $\psi(y)$ is real-valued, measurable, and invariantly predictable, (iii) (8.1) has a unique solution for $\overline{\psi}(x,\gamma)$ for all $x\in\mathcal{X}$, $0 < \gamma < 1$, and (iv) $g'\psi$ increases as ψ increases for each $g'\in\mathcal{G}'$, then (8.2) is satisfied.

Proof. The defining equation for $\overline{\psi}(x,\gamma)$ is

$$(8.3)\qquad \Delta(t)\int_{\psi(y)\leq\overline{\psi}(x,\gamma)}\left\{\int_\Omega f(\omega^{-1}t,\omega^{-1}y|a)J(\omega)\nu(d\omega)\right\}\xi(dy) = \gamma.$$

Substituting gx = (gt,a) for x = (t,a) gives

$$(8.4)\qquad \Delta(gt)\int_{\psi(y)\leq\overline{\psi}(gx,\gamma)}\left\{\int_\Omega f(\omega^{-1}gt,\omega^{-1}y|a)J(\omega)\nu(d\omega)\right\}\xi(dy) = \gamma.$$

Changing to new integration variables $\omega_1 = g^{-1}\omega$, $y_1 = g^{-1}y$, and using $\Delta(gt) = \Delta(g)\Delta(t)$, $\Delta(g)\nu(d\omega) = \nu(d\omega_1)$, $J(g\omega_1) = J(g)J(\omega_1)$, and $J(g)\xi(dy) = \xi(dy_1)$ gives

$$(8.5)\qquad \Delta(t)\int_{\psi(gy_1)\leq\overline{\psi}(gx,\gamma)}\left\{\int_\Omega f(\omega_1^{-1}t,\omega_1^{-1}y_1|a)J(\omega_1)\nu(d\omega_1)\right\}\xi(dy_1) = \gamma.$$

Since $\psi(y)$ is invariantly predictable, we may write $g'\psi(y_1)$ in place of $\psi(gy_1)$, and by assumption (iv), the y_1 integration in (8.5) is over values such that $\psi(y_1) \leq g'^{-1}\overline{\Psi}(gx,\gamma)$. Comparison of (8.5) and (8.3) then gives, by (iii), $\overline{\Psi}(x,\gamma) = g'^{-1}\overline{\Psi}(gx,\gamma)$. If we let $I(x,y)$ denote the indicator function which equals 1 or 0 according as $\psi(y)$ is $\leq$ or $>$ $\overline{\Psi}(x,\gamma)$, it follows that $I(gx,gy) = I(x,y)$. Appealing to Theorem 5.1 we have

$$
\begin{aligned}
P\{\psi(y) \leq \overline{\Psi}(x,\gamma)\,|\,\omega\} &= E^{x,y|\omega} I(x,y) \\
&= E^{x,y|\omega} E^{y,t|a,\omega} I(x,y) \\
&= E^{x,y|\omega} E_f^{y,\omega|x} I(x,y) \\
&= E^{x,y|\omega} P_f\{\psi(y) \leq \overline{\Psi}(x,\gamma)\,|\,x\} \\
&= E^{x,y|\omega}\gamma = \gamma .
\end{aligned}
$$

From the equivalence of fiducial and posterior distributions noted in Section 4 we have:

Corollary 8.1. Under the conditions of the theorem, Bayesian limits for $\psi(y)$, based on prior measure ν, have the frequency interpretation (8.2).

A result announced by Hall and Novik (1963) is more general than Corollary 8.1 in that it includes a regression parameter, but more special in being restricted to certain location and scale parameter models.

9. Remarks on some consistency criteria for fiducial distributions. Fisher (1956) states, "The concept of probability involved in the fiducial argument is entirely identical with the classical probability of the early writers, such as Bayes," (p 51), and later on p 125, "This fiducial distribution supplies information of exactly the same sort as would a distribution *a priori*." Lindley (1958) put these assertions to test by calculating in three different ways the fiducial distribution of a parameter θ given two observations x_1 and x_2: (1) from a sufficient statistic depending on (x_1,x_2), (2) from the posterior distribution of θ given x_2 when the prior distribution of θ is equated to the fiducial distribution of θ given x_1, (3) same as (2) with x_1 and x_2 reversed. Consistency in Lindley's sense is the equality of the three resulting distributions of θ given (x_1,x_2). It is clear from the consistency of the Bayesian method that equality holds whenever the fiducial distributions are posterior distributions corresponding to the same prior. Since this is known to be the case when Assumptions 1 through 5 above are satisfied (with prior measure equal to right Haar measure--see for example HB-1, Section 2.2),Lindley's consistency (in a generalized form) holds in the present circumstances.

In Section 4 and in the Appendix we have indicated how the fiducial distribution (4.6) of future observations y can be obtained either (1) by Bayesian analysis, or (2) by a pivotal argument. Still another derivation, which we may call the "integral method", obtains (4.6) by multiplying the density of y given x and ω by the fiducial density of ω given x and then integrating over ω. Again the equivalence of the Bayesian and fiducial distributions guarantees agreement. The integral method was used by Sprott (1963) and seems to be implicit in Fisher (1956), p 126. Note that here the fiducial distribution of ω given x is being used as a prior distribution, as in Lindley (1958), but in a different way. The agreement just mentioned is thus another consistency property of fiducial distributions having the form (4.5). A somewhat more stringent criterion proposed by Buehler (1963) requires that prediction limits for $\psi(y)$ obtained by the integral method should have the frequency property (8.2). Theorem 8.2 shows that in the present framework, invariant predictability of $\psi(y)$ (with mild regularity conditions) is a sufficient condition, and generalizes the results announced by Buehler (1963).

APPENDIX

<u>Derivation of the fiducial distribution of y given x by the "pivotal" method</u>. We begin with a simple case to illustrate the type of argument. Suppose $f(x-\theta,y-\theta)$ is a bivariate density, and it is desired to obtain the fiducial distribution of y given x (not to be confused with the (nonfiducial) distribution of y given x <u>and</u> θ). We first find the marginal distribution of $v = x-y$ given θ and show that it does not actually depend on θ. Thus v is an ancillary statistic, and a suitable "pivotal" for our purposes. The transformation from v to y with x fixed yields a distribution not depending on θ, the fiducial distribution of y given x. We now proceed with a similar analysis for the general case.

From (2.3), (2.4) and (4.3) we obtain the joint distribution of (t,y) given (a,ω) as

$$f(\omega^{-1}t,\omega^{-1}y|a)\mu(dt)J(\omega)\xi(dy). \tag{A.1}$$

Defining $v = t^{-1}y$ we obtain the joint distribution of (t,v) given (a,ω) as

$$f(\omega^{-1}t,\omega^{-1}tv|a)\mu(dt)J(t^{-1}\omega)\xi(dv). \tag{A.2}$$

To obtain the marginal distribution of v given (a,ω), we integrate the last expression with respect to t, and after changing the integration variable to $z = \omega^{-1}t$ this yields

$$\xi(dv)\int f(z,zv|a)J(z^{-1})\mu(dz). \tag{A.3}$$

In this form it is clear that the distribution of v given a does not depend on ω, so that v is a conditional pivotal quantity. Next consider t to be fixed, and transform the variate v to y where $v = t^{-1}y$. This "pivotal argument" yields the fiducial distribution of y given $x = (t,a)$ in the form

$$\xi(dy)J(t)\int f(z,zt^{-1}y|a)J(z^{-1})\mu(dz). \tag{A.4}$$

On changing the integration variate from z to ω where $z = \omega^{-1}t$, the last expression will be seen to agree with (4.6). Thus the "pivotal argument" leads to the same result as the Bayesian analysis.

REFERENCES

Barnard, G.A. (1952). The frequency justification of certain sequential tests. Biometrika **39** 144-150.

Buehler, R.J. (1963). A new test of fiducial consistency (Abstract). Ann. Math. Statist. **34** 1112.

Eisenhart, C. (1939). The interpretation of certain regression methods and their use in biological and industrial research. Ann. Math. Statist. **10** 162-186.

Fisher, R.A. (1935). The fiducial argument in statistical inference. Annals of Eugenics **6** 391-398.

Fisher, R.A. (1956). Statistical Methods and Scientific Inference. Oliver and Boyd, Edinburgh.

Fraser, D.A.S. (1961). The fiducial method and invariance. Biometrika **48** 261-280.

Hall, W.J. and Novick, M.R. (1963). A note on classical and Bayesian prediction intervals for location, scale and regression models (Abstract). Ann. Math. Statist. **34** 1619.

Hartigan, J. (1964). Invariant prior distributions. Ann. Math. Statist. **35** 836-845.

Hora, R.B. and Buehler, R.J. (1965). Fiducial theory and invariant estimation. To appear in June Ann. Math. Statist. (Published in 1965 as University of Minnesota Department of Statistics Technical Report No. 42 (revised).)

Jeffreys, H. (1960). Theory of Probability (3rd ed.). Oxford University Press.

Kitagawa, T. (1951). Successive process of statistical inferences (2). Memoirs of the Faculty of Science, Kyusyu University Ser. A **6** 55-95.

Kitagawa, T. (1957). Successive process of statistical inference associated with an additive family of sufficient statistics. Bulletin of Mathematical Statistics (Fukuoka) 7 92-112.

Kudo, A. (1956). On the confidence interval of the extreme value of a second sample from a normal universe. Bull. Math. Stat. **6** 51-56.

Lindley, D.V. (1958). Fiducial distributions and Bayes theorem. J. Roy. Statist. Soc. Ser. B **20** 102-107.

Mood, A.M. (1950). Introduction of the Theory of Statistics. McGraw-Hill, New York.

Peisakoff, M.P. (1950). Transformation parameters. Ph.D. Thesis, Princeton University.

Ramsey, F.L. (1963). Some fiducial prediction problems. MS Thesis, Iowa State University, Ames, Iowa.

Roy, A.D. (1960). Some notes on pistimetric inference. J. Roy. Statist. Soc. Ser. B **22** 338-347.

Thatcher, A.R. (1964). Relationships between Bayesian and confidence limits for predictions. J. Roy. Statist. Soc. Ser. B **26** 176-192.

Weiss, L. (1955). A note on confidence sets for random variables. Ann. Math. Statist. **26** 142-144.